

Tartu Ülikool  
Sotsiaalteaduste valdkond  
Psühholoogia instituut

Jane Kuperjanov

Lapse sõnavara ja tema keelekeskkond: LENA kõneanalüüsi süsteemi kasutamine eesti laste  
keelekeskkonna mõõtmisel

Magistritöö

Juhendajad: Tiia Tulviste (PhD), Ada Urm (MA)

Läbiv pealkiri: LENA kõneanalüüsi süsteemi kasutamine eesti laste keelekeskkonna  
mõõtmisel

Tartu 2019

## **Lapse sõnavara ja tema keelekeskkond: LENA kõneanalüüsi süsteemi kasutamine eesti laste keelekeskkonna mõõtmisel**

### **Kokkuvõte**

Käesoleva töö peamiseks eesmärgiks oli uurida LENA kõneanalüüsi süsteemi usaldusväärsust eesti keelekeskkonnas. Lisaks uuriti, kas lapsed, kelle kodune keelekeskkond LENA kõneanalüüsi süsteemi andmetele tuginedes on rikkam, saavad paremaid tulemusi ka ECDI–II testiga. Eraldi eesmärgiks oli uurida vanemate haridustaseme ja laste soo seoseid ECDI–II testi tulemustega. Tulemustest selgus, et LENA kõneanalüüsi süsteemi saab eesti laste keelekeskkonna uurimisel kasutada. LENA süsteemiga saadud laste vokalisatsioonide, täiskasvanute sõnade, laste voorude arv oli sarnane CLAN programmiga analüüsitud 10–minutiliste transkribeeritud lõikude tulemustega. Laste vokalisatsioonide, täiskasvanute sõnade ja laste voorude ning ECDI–II testi tulemuste vahel olid positiivsed, ent statistiliselt mitteolulised seosed. Tugev positiivne seos ilmnes poiste valimi puhul ECDI–II testi tulemuste ja täiskasvanute sõnade vahel, mis toetab varasemaid tulemusi keelesisendi olulisusest laste kõne arengus. Vanemate hariduse ja laste ECDI–II testi tulemuste vahel arvestatavad seosed puudusid.

*Märksõnad:* LENA, kõne areng, ekspressiivne sõnavara, täiskasvanute sõnad, lapse vokalisatsioonid, lapse voorud, sotsiaalmajanduslik staatus, sõnavara suurus

## **Child's vocabulary and language environment: Using the LENA System to measure the language environment of Estonian children**

### **Abstract**

The main aim of this thesis was to study the reliability of the LENA System in the Estonian language environment. In addition, it was investigated whether children whose home language environment is more complex based on the LENA System, will get better results with the ECDI–II test. The results of this test were also compared with the parents' educational level and the child's gender and studied for possible associations.

The results proved that the LENA System can be used for research into the language environment of Estonian children. The number of child vocalisations, adult words and child turns determined with the LENA System was similar to the results of the 10-minute transcribed sections analysed with the CLAN programme. When comparing a general selection of child vocalisations, adult words and child turns with the ECDI–II test results, the results were positive but insignificant. A strong association was observed among the boys' ECDI–II test results and adult words, which supports earlier findings on the importance of language input in the development of children's vocabulary. There were no significant connections between parents' education and ECDI–II test results.

*Keywords:* LENA, language development, expressive vocabulary, adult words, child vocalisations, child turns, socio-economic status, vocabulary size

## Sissejuhatus

LENA (*the Language ENvironment Analysis System*) on keelekeskkonna uurimise süsteem, mis koosneb spetsiaalsest tarkvarast ja salvestusseadmest, võimaldades salvestada kõnet loomulikus keskkonnas ning salvestisi efektiivselt analüüsida [[www.lena.org](http://www.lena.org)]. Käesolev töö on esimene katsetus LENA kõneanalüüsi süsteemi usaldusväärsuse uurimiseks eesti keelekeskkonnas. Lisaks uuriti töö raames LENA seadmega kogutud andmete põhjal laste keelekeskkonna seost nende aktiivse sõnavaraga.

Laste keele arengu uurimine võimaldab individuaalsete erinevuste selgitamise kaudu täiendada teaduse teooriaid ning kavandada tõhusaid, tõenduspõhiseid sekkumisi kõne arengu puudujääkide korral (Golinkoff, Hoff, Rowe, Tamis-LeMonda & Hirsh-Pasek, 2018). Kõneprobleemide põhjuste väljaselgitamine on oluline. Vahel võib lapse kõne arengu mahajäämus olla tingitud kuulmis- või nägemisprobleemidest, kõne arengu peatumine võib omakorda põhjustada käitumuslikke, emotsionaalseid ja sotsiaalseid raskusi. Võimalikult varajane märkamine ja sekkumine võimaldab vähendada tõsisemate kahjustuste tekkimist.

### Lapse– ja täiskasvanu kõne seosed

Käesoleva töö raames uuriti lapsi, kelle vanus jäi uuringuperioodil 34 ja 42 elukuu vahele. Selles vanuses on tavaarenguga laste puhul kõne arengu varieeruvus suur, kuid üldiselt kasutavad 3–aastased lapsed 900–1000 erinevat sõna ja 3–4 sõnalisi lauseid (Nicholas & Geers, 2006). Paavola, Kunnari ja Moilanen (2005) leidsid oma uuringus, et kui lapsel on rohkem vokalisatsioone kõne-eelsel perioodil, on nad hiljem kõne arengus kiiremad. Laste keele arengut mõjutavad nii geneetilised kui ka keskkondlikud tegurid (Stromswold, 2001) ning on leitud, et ekspressiivse kõne areng on suures osas keskkonnateguritega seotud (VanHulle, Goldsmith & Lemery, 2004). Hart ja Risley (1995) tõid oma pikaajalises sotsiaalmajandusliku staatuse ja kõne arengu seoste uuringus välja suured erinevused perede vahel selles osas, kui palju lastega räägitakse ning kui mitmekesisest sõnavara kasutatakse. Nad leidsid, et vanemate kasutatud sõnavara hulk ja mitmekesisus olid positiivselt seotud laste sõnavara kasvukiirusega, samuti leiti, et kõrgema sotsiaalmajandusliku staatusega peredes kuulsid lapsed kõnet rohkem (Hart & Risley, 2003). Pere sotsiaalmajandusliku taseme hindamise tegurid on vanemate haridus, elukutse ja sissetulek, ent kõige olulisemaks peetakse hariduse taset (Westerlund & Lagerberg, 2007).

Keele omandamiseks vajab laps keelemudelit, suhtluspartnerit ja suhtlemiskogemust. Kui laps on jäetud üksi aega veetma vaid televiisoriga, puudub keelemudel ning suhtluspartner (Hoff, 2006). Kõige enam näib head kõne arengut soosivat see, kui lastega räägitakse neile huvipakkuvatel teemadel ja neid kaasatakse vestlusesse. Lapsed, kes kuulevad rohkem neile suunatud kõnet, pikemaid lauseid, mis sisaldavad rikkamat sõnavara, on süntaktilises arengus paremal tasemel ning neil on täheldatud sõnavara kiiremat suurenemist (Hoff, 2003). Uuringutega on samuti tõestust leidnud, et kõrgharidusega emad vestlesid rohkem oma 2–aastaste lastega, kasutasid rikkamat sõnavara, vastasid sagedamini laste küsimustele, kasutasid vähem direktiivset suhtlemisstiili ja küsisid rohkem küsimusi, kui keskkooli lõpetanud emad (Hoff-Ginsberg, 1998). Schults, Tulviste ja Konstabel (2012) töid oma töös välja positiivsed seosed lapse sõnavara suuruse ja ema hariduse vahel.

Hilisemateaju–uuringutega on kinnitust leidnud, et väikelapse poolt kuuldav keele hulk ja mitmekesisus ning lausete keerukus on hilisema keele- ja kirjaoskuse alus aga ka eksekutiivsete funktsioonide, matemaatiliste- ja sotsiaalsete oskuste mõjutaja (Romeo jt, 2018). Romeo ja kaastöötajate (2018) uuringus leiti, et just vanematega vesteldes aktiveerus laste ajus Broca piirkond, samuti ilmnis seos vanemate hariduse ja laste keeleoskuse skooride vahel.

Laste arengut jälgides ja hinnates on oluline mitte unustada keele arengu soolist eripära. Tüdrukute kõne areng kaldub olema poiste omast kiirem (Yu jt, 2014), nad edestavad poisse produktiivse sõnavara, verbaalse kõne mõistmise (Berghlund, Eriksson & Westerlund, 2005) ja kommunikatiivsete žestide arengu osas (Eriksson jt, 2012). Samuti esineb poistel kõnehäireid statistiliselt sagedamini kui tüdrukutel (Campbell jt, 2003; Choudhury & Benasich, 2003).

Enamasti on uuringutes rõhutatud lapsele suunatud kõne ning selle kvaliteedi olulisust (Golinkoff jt, 2018). Sperry, Sperry ja Miller (2018) viitasid sellele, et vähem on uuritud kõnet, mida lapsed kuulevad, kuid mis ei ole suunatud otseselt neile (*overheard speech*), ehkki nad sellise kõnega samuti igapäevaselt kokku puutuvad ning mis seetõttu laste kõne arengu mõjutajana oluline on. Eelnevates uurimustes on näidatud seda, et lapse kõne arengu kõige olulisem mõjutaja on hooldaja silmast silma diaadiline suhtlemine lapsega, lapse vestlusesse kaasahaaramine, hooldajapoolne sõnade kirjeldamine ja kordamine (Hoff, 2003; Hoff-Ginsberg, 1998; Golinkoff jt, 2018).

### **Varasemad uuringud LENA kõneanalüüsi süsteemiga**

Kõne arengu uuringud, mis toetuvad lapse igapäevase suhtluse analüüsile, on väga ressursimahukad. Salvestuste kogumine on aeganõudev, samuti salvestatud teksti

transkribeerimine ning analüüsimine. Lisaks võib lõpphinnangu valiidsust mõjutada uuringu koht – kui laps viibib uuringus osalemiseks laboris, võib see mõjutada lapse kõne produtseerimist uurimise hetkel, kuna laps võib olla häiritud ja ei käitu tavapäraselt.

Kõneuringute ressursimahukuse ja kontekstist tulenevate moonutuste leevendamiseks löid USA uurijad Terrance ja Judith Paul Ameerika Ühendriikidest Hart'i ja Risley uuringust inspireerituna LENA kõneanalüüsi süsteemi. Kaasaskantav elektrooniline LENA seade võimaldab last uurida suhteliselt häirimatult lapse ja vanema jaoks loomulikus keskkonnas [<https://www.lena.org/>].

LENA kõneanalüüsi süsteem koosneb audiosalvestusseadmest, töötlemisprogrammist ja lastele mõeldud taskuga t-särgist või vestist, kuhu seade paigutatakse. Süsteem on mõeldud 0–4. aastaste laste keele arengu uurimiseks ja sekkumisvajaduse väljaselgitamiseks.

Inglisekeelse valimi alusel on välja selgitatud normid uuritava 2–48 kuu vanuse lapse arengu võrdlemiseks eakaaslastega (Gilkerson & Richards, 2008). LENA kõneanalüüsi süsteem on hinnatud usaldusväärseks täiskasvanute sõnade puhul, mida seade tuvastas 82% ulatuses ja kuni 3-aastaste väikelaste sõnade puhul, mida tuvastati 76% ulatuses (Gilkerson jt, 2015; Zimmerman jt, 2009).

Enamik LENA süsteemi uuringutest on läbi viidud tavaarenguga lastega, kuid uuritud on ka autismispektrihäiretega-, kuulmiskahjustustega- ja Down'i sündroomiga lapsi, enneaegselt sündinud imikuid ning keele arengu pidurdumisega lapsi (Ganek & Eriks-Brophy, 2016). Lisaks on soome uurija Hanna Elo (2016) LENA süsteemi abil uurinud kaksikute keele arengut.

LENA uuringuid on läbi viidud väga erinevas vanuses lastega. Näiteks prantsuse uuringus on osalejate vanusevahemik 3–48 kuud (Canault, Normand, Foudil, Loundon & Thai-Van, 2015), Hollandis uuriti (Busch, Sangen, Vanpoucke & Wieringen, 2017) aga 2–5 aasta vanuseid lapsi, mida tuleb arvestada uuringutulemuste võrdlemisel.

Juba terve dekaad on LENA kõneanalüüsi süsteemi kasutatud uuringutes ingliskeelses keskkonnas. Ent süsteemi usaldusväärsust on hakatud testima ja LENA abil uuringuid läbi viima ka teistes keelekeskkondades, näiteks prantsuse (Canault jt, 2015), hispaania (Weisleder & Fernald, 2013), hiina (Gilkerson jt, 2015), soome (Elo, 2016), korea (Pae jt, 2016), rootsi (Schwarz jt, 2017), hollandi (Busch jt, 2017) ja vietnami keeles rääkivate laste keelekeskkonna uurimisel (Ganek & Eriks-Brophy, 2018).

Kõik teistes keeltes LENA kõneanalüüsi süsteemiga läbi viidud uurimused ei ole keskendunud samadele keeleaspektidele. Mõnes töös on uuritud süsteemi usaldusväärsust näiteks ainult

täiskasvanute sõnade kokkulugemisel (Schwarz jt, 2017), teistes on püütud välja selgitada seadme usaldusväärsus kõigis neljas segmendis (laste vokalisatsioonid, täiskasvanu sõnad, lapse vooruvahetused, elektrooniline taust, nt TV jm) (Busch jt, 2017).

*Uuringud laste vokalisatsioonide kohta LENA kõneanalüüsi süsteemiga.* Laste vokalisatsioonide tuvastamist ja võrdlust inimese poolt tehtud transkriptsioonide analüüsimisel saadud tulemustega uuriti hiina, hollandi, prantsuse ja soome uuringutes.

Hollandi keelekeskkonnas salvestatud laste vokalisatsioonide puhul ilmnas tugev korrelatsioon transkriptsioonide ja LENA süsteemi loenduste vahel ( $r=.77$ ; Busch jt, 2017), vaid veidi nõrgemaks ( $r=.71$ ,  $p < .001$ ) jäi vastav seos prantsuse tööühma poolt saadud tulemuste puhul (Canault jt, 2015).

Ka Soome kaksikute uuringus tuvastati inimese transkriptsioone ja LENA kõneanalüüsi süsteemi kodeeringut võrreldes süsteemi väga kõrge reliaablus – esimese kaksiku puhul (*key child*) 84%, teise kaksiku puhul (*other child*) 96% ja LENA automaatsed arvutused lapse vokalisatsioonide hulga kohta osutusid usaldusväärseteks (Elo, 2016).

*Täiskasvanute sõnade arvu* usaldusväärsust LENA loendustes on uuritud rootsi, hollandi, prantsuse, hispaania ja soome keelte uuringutes.

Schwarzi tööühma rootsi keelekeskkonna uuringus (2017) leiti LENA kõneanalüüsi süsteemi täiskasvanute sõnade hinnangu ja transkribeerijate sõnade arvu vahel mõõdukas korrelatsioon ( $r=.67$ ,  $p < .05$ ). Seega osutusid LENA süsteemi hinnangud täiskasvanute sõnade arvu puhul rootsi keeles kõige sarnasemaks prantsuse keelega (Schwarz jt, 2017).

Prantsuse uuringu (Canault jt, 2015) puhul leiti, et LENA kõneanalüüsi süsteemi poolt loendatud täiskasvanute sõnade arv oli korrelatsioonis transkribeerijate hinnangutega ( $r=.64$ ,  $p < .001$ ).

Inimese poolt loendatud täiskasvanute sõnade skoor oli suurem kui LENA täiskasvanute sõnade skoor prantsuse keele uuringu puhul (110.318 vs 73.274 sõna), nende kahe hinnangu keskmine suhe oli 1,56 (Canault jt, 2015).

Hispaania keelekeskkonna uuringus ilmnas, et täiskasvanute sõnade LENA automaatse loendi ja transkriptsioonil põhinevate sõnade vahel suur korrelatsioon ( $r=.80$ ), osutades LENA kõneanalüüsi süsteemi usaldusväärsusele täiskasvanute sõnade lugemisel (Weisleder & Fernald, 2013).

Soome kaksikute uuringus ilmnas inimese ja LENA kodeeringu võrdluse tulemustel naissoost täiskasvanu puhul väga hea (85% ja 89%), kuid meessoost täiskasvanu puhul mõõdukas (49% ja 77%) reliaablus (Elo, 2016).

*Voorude uuringud.* Ganek ja Eriks-Brophy uurisid oma 2018 aastal avaldatud töös LENA seadme usaldusväärsuse hindamiseks laste vooruvahetusi Spearmani korrelatsioonitestiga, mis näitas tugevat korrelatsiooni LENA tarkvara ja transkriptsioonide vahel ( $r(18)=.70$ ,  $p<.001$ ) ning leidsid, et LENA süsteem annab vietnami keele puhul vestlusvoorude kohta suhteliselt täpse hinnangu (Ganek & Eriks-Brophy, 2018). Hollandi keele uuringus (Busch jt, 2017) tulid korrelatsioonid vestlusvoorude puhul aga madalamad ( $r=.52$ ), eriti võrreldes Hiina uuringuga ( $r=.72$ ; Gilkerson jt, 2015).

### **Varasemad uuringud, kus on kasutatud LENA kõneanalüüsi süsteemi ning arenguteste**

Käesolevas uuringus on kasutatud MacArthuri-Batesi suhtlemise arengu testi „Sõnad ja laused“ eesti adaptatsiooni (ECDI–II) (Tulviste, 2007), et võrrelda LENA loendatud sõnade ja kõnevoorude hulka ning arengutesti tulemusi. ECDI–II ja LENA uurivad laste keele arengu erinevaid aspekte, esimene aktiivset sõnavara ning grammatikareeglite valdamist, teine sõnade ja voorude hulka. Kuna on leitud, et vanema suhtlemine lapsega on alus lapse sõnavara kasvukiirusele, hilisemale akadeemilisele edukusele ja lapse üldisele arengule, oli töö eesmärgiks otsida vastust küsimusele, kas paremad tulemused LENA kõneanalüüsi süsteemiga ennustavad paremat tulemust ECDI testi puhul ja vastupidi. Kuigi ECDI–II test ja LENA hindavad erinevaid keele arengu aspekte, on siiski LENA kõneanalüüsi süsteemi reliaabluse hindamiseks oluline, et mõlema meetodi puhul on tulemused sarnased.

Varem on MB–CDI (MacArthur–Bates Communicative Development Inventory) testi ja LENA seadet kasutanud oma uuringus Warren jt (2010). Warreni uuringurühm uuris keelekeskkonna erinevusi autismispektri häiretega (ASD) laste peredes ja võrdlesid tulemusi tavaarenguga laste tulemustega. Uuriti vanemate kõne hulka, laste vokalisatsioonide hulka ja vestlusvoorude hulka ning võrreldi kontrolligrupiga. Lapsed olid vanuses 16–48 kuud. Salvestusteks kasutati LENA seadet ning vanemad täitsid 9 küsimustikku oma lapse arengu kohta, muuhulgas MB–CDI. Kõige tugevam positiivne seos esines ASD laste puhul MB–CDI sõnavara suuruse ja LENA vestlusvoorude vahel. Mida rohkem loendati täiskasvanute sõnu ja vooruvahetusi, seda väiksem oli häire sümptomaatika ja suurem keeleline võimekus testide järgi (Warren jt, 2010). Elo (2016) kasutas laste sõnavara suuruse ja varajase keeleoskuse arengu uurimiseks samuti muuhulgas MB–CDI testi soomekeelset verisooni. Kui võrreldi LENA üldisi tulemeid MB–CDI skooridega (kasutades Spearmani mitteparameetrilist korrelatsioonianalüüsi), ilmnas



ainsana mõõdukas positiivne seos täiskasvanute keskmise sõnade arvu ning laste suhtlusmeetmete ja žestide kasutamise mitmekülgsuse vahel. Käsitletavas uuringus ei leitud rohkem statistiliselt olulisi seoseid laste vokalisatsioonide ja voorude ning MB–CDI keele skooride vahel. Weisleder ja Fernald (2013) kasutasid 19–24 kuu vanuseid lapsi hõlmanud uuringus MacArthur–Batesi testi (MB–CDI) hispaaniakeelset versiooni. Nad kasutasid LENA hinnanguid, MB–CDI testi hispaaniakeelset versiooni 24–ndal elukuul ekspressiivse sõnavara mõõtmiseks ning vaatamise ja kuulamise ülesannet keele töötlemise efektiivsuse mõõtmiseks. Uurijad leidsid, et lapsed, kes kuulsid 19 kuu vanusena vanemalt lapsele suunatud kõnet rohkem, omasid 5 kuu pärast suuremat sõnavara. Lapsele lihtsalt kuuldav kõne (*overhead speech*) ei olnud seotud lapse sõnavara suurusega ja autorid järeldasid, et just lapsele suunatud kõne toetab sõnavara arengut (Weisleder & Fernald, 2013). Ülaltoodud uuringutes on leitud tugevamaid (Warren jt, 2010) ning nõrgemaid seoseid (Elo, 2016) CDI arengutesti ja LENA poolt loendatud sõnade ja voorude vahel ning seetõttu oli huvipakkuv välja selgitada vastavad tulemused minu poolt uuritava valimil.

Varasemates töödes oli LENA usaldusväärsuse väljaselgitamiseks transkribeeritud erineva pikkusega salvestiste lõike, osades uuringutes leiti ka transkribeerijate usaldusväärsus. Näiteks Canault jt töörühm (2015) valis prantsuse keelekeskkonna uuringus iga katseisiku (N=18) salvestusest välja kuus 10–minutilist salvestiste lõiku, s.o 324 lõiku (kokku 54 tundi). 108 lõiku transkribeeriti kahe prantsuse keelt kõneleva eksperdi poolt, kelle transkribeeringute vahel oli tugev korrelatsioon ( $r=.99$ ,  $p<.001$ ). Elo (2016) uuringus koguti LENA kõneanalüüsi süsteemiga andmeid 6–12 kuu vanuste laste kohta 2 päeva jooksul ning lisaks LENA seadme usaldusväärsuse hindamisele teostati ka transkribeerijate usaldusväärsuse hindamine. Selleks transkribeeriti kahe transkribeerija poolt salvestuste kogumahust lõikudena 10% (132 minutit) sarnaste juhiste alusel.

### **Magistritöö eesmärk ja uurimisküsimused**

Toetudes eelnevatele uuringutele seadsin töö eesmärkideks:

- töö esimeses osas uurida, kuidas LENA kõneanalüüsi süsteem loendab täiskasvanute sõnu, laste vokalisatsioone ja laste voore eesti keeles võrreldes transkribeerija poolt saadud tulemustega.
- töö teises osas uurida, kuidas LENA seadmega ja lastevanemate raportitega (ECDI–II) kogutud andmed on omavahel seotud ning välja selgitada, kas on olulisi erinevusi vanemate haridustaseme ja laste soo lõikes ECDI–II testi tulemustele uuritava valimil.

## Uurimisküsimused

Uuringu eesmärgiks oli otsida vastuseid küsimustele:

1. Kas LENA meetod on eesti keeleruumis usaldusväärne laste vokalisatsioonide, laste voorude ja vanemate sõnade loendamiseks?
  - 1.1. Millised on kahe meetodi (LENA vs inimese transkriptsioonid) erinevused?
2. Kas ECDI–II testi tulemused ja LENA loendatav laste vokalisatsioonide, vooruvahetuste ning täiskasvanute sõnade arv on omavahel positiivselt seotud koondvalimil ja eraldi poiste ning tüdrukute alavalimil.
  - 2.1. Kui palju kõnet kuulsid uuritavad lapsed täiskasvanutelt ja kui palju rääkisid lapsed ise ühe salvestuspäeva jooksul?
  - 2.2. Kuidas on seotud laste vanemate haridustase laste ECDI–II testi tulemustega uuritaval valimil?

## Meetod

### Valim

Käesolevas uuringus analüüsiti longituuduuringu „Varase kasvukeskkonna toetav roll lapse kõneoskuste arengus: võimalus kahandada lõhet hilisemates akadeemilistes saavutustes“ (PUT 1359, vastutav täitja Tiia Tulviste) esimese laine andmeid. Luba longituuduuringu läbiviimiseks saadi TÜ Inimuuringute komiteelt. Longituuduuringusse kaasati Tartu piirkonna lasteaedade ~3 aastased lapsed vanemate nõusoleku korral. LENA uuringus paluti osaleda juhusliku valiku alusel 28 lapse vanemal, kellele edastati LENA salvestusseade. Kaks last ei kandnud seadet, 3 salvestust osutusid tehnilistel põhjustel tühjaks ning ühe lapse puhul ei olnud kogu sõnade hulk salvestunud, kuna seade ei asetsenud terve päeva jooksul lapsele piisavalt lähedal. Seega kujunes lõplikuks valimiks 22 last (13 tüdrukut ja 9 poissi).

Longituuduuringu esimesse lainesse kaasatud lapsed olid ECDI-II testi täitmise ajal 34–38 kuu vanused ( $M=35,7$ ;  $SD=0,9$ ). Andmed vanuse kohta testi täitmise ajal olid olemas 20 osaleja kohta.

Salvestus LENA seadmega viidi läbi veidi hiljem ja lapsed olid sel ajal 35–42 kuu vanused ( $M=37,8$ ;  $SD=1,7$ ), need andmed olid olemas kõigi 22 lapse kohta.

Vanemad täitsid küsimustikus andmed nii ema kui ka isa hariduse kohta (Tabel 1). Ühegi lapse kohta ei olnud vanemad Kaemuse testikeskkonnas välja toonud terviseprobleeme ning kõikide laste kodudes räägiti vanemate kinnitusel eesti keelt emakeelena.

**Tabel 1.** Uuritavate laste vanemate jaotus haridustaseme kaupa

	Ema haridus		Isa haridus	
	N	Osakaal	N	Osakaal
Kõrgharidus	16	73%	10	48%
Kutseharidus	2	9%	8	37%
Keskharidus	4	19%	2	10%
Põhiharidus	0	—	1	5%

### Protseduur

Longituuduuringu esimeses laines paluti lapsevanematel täita TÜ Psühholoogia instituudi testimiskeskonnas Kaemus taustaandmete ankeet, milles paluti märkida lapse sugu, sünniaeg, lapse tervislik seisund, mõlema lapsevanema haridustase jm. Lapsevanema poolt täidetud taustaandmed olid käesoleva töö puhul aluseks valimi kirjeldamisel ja analüüside tegemisel.

Lisaks täitsid lapsevanemad testimiskeskonnas Kaemus oma lapse kõne arengu kohta ECDI–II testi.

Uurimistöös kasutati laste keelekeskkonna uurimiseks innovaatilist LENA kõneanalüüsi süsteemi. Perelele anti väike ja turvaline LENA salvestusseade, mida laps kandis seadeldisega koos antava vesti või t-särgi taskus ühel nädalavahetuse päeval ärkamisest kuni magama minekuni. Lisaks täitsid vanemad uuringus osalemise nõusolekulehed ning keelekeskkonna päeviku, kuhu salvestuse päeva kohta tuli 1–tunniste intervallidega kirjutada, kellega koos laps oli ning kus viibiti (kodus, õues jne). Samuti sai päevikus ära märkida ajavahemiku, mille pere soovis analüüsist välja jätta. Seda soovi arvestati transkribeerimislõikude väljavalimisel ja neid lõike ei kasutatud. Kuna uurijat salvestamise ajal juures ei olnud ning väikelapse kõnet on inimesel, kes lapsega pidevalt kokku ei puutu, vahel raske mõista, oli päevik abiks transkribeerijale vestlusesse kaasatud isikute kindlakstegemiseks (täiskasvanu või laps) ja jutu konteksti mõistmiseks.

Salvestused LENA seadmega viidi läbi ajavahemikul 20.10.2017 – 04.07.2018.

### **Transkribeerimine**

LENA seadmega kogutud salvestiste kogupikkus oli 329 tundi ( $M=14:57:27$ ;  $SD=1:07:01$ ). Kuna uurimistöö üheks eesmärgiks oli uurida LENA süsteemi abil saadavaid tulemusi ning võrrelda neid ECDI–II testi tulemustega, otsustati keskenduda sellele, kui täpselt LENA süsteem eristab ja loeb lapse poolt produtseeritud eestikeelset kõnet. Igast salvestusest valiti kaks 5-minutilist järjestikust lõiku, kus laps rääkis päeva jooksul kõige rohkem. 230 minutit transkribeeriti CLAN programmi kasutades autori poolt, kes räägib eesti keelt emakeelena. 10% nendest välja võetud lõikudest transkribeeriti usaldusväärse väljaselgitamiseks ka teise eesti keelt emakeelena kõneleja poolt, sarnaselt Elo (2016) tööga.

Transkriptsioone analüüsiti CHILDES süsteemi kokkuleppeid kasutades (Child Language Data Exchange System) ning analüüsides eemaldati vestluses osalejate kattuv kõne.

### **Mõõtmisvahendid**

#### ***LENA kõneanalüüsi süsteem***

Käesolevas töös kasutati laste keelekeskkonna uurimiseks LENA Pro tudengi versiooni, mis võimaldab uurida korduvalt kuni 25 katseisikut ja sisaldab seadet, riidetust ning tarkvara. LENA seade on pisike, kaalub umbes 50 grammi ning võimaldab salvestada 16 tundi. Seadet on kerge käsitseda. Piisab sisselülitamisest ning salvestama panemisest, 16 tunni möödudes lülitab seade

end ise välja. Salvestus laaditakse pärast lindistamist LENA tarkvaraga varustatud arvutisse. Pärast salvestuse mahalaadimist saab seadet uuesti kasutada.

LENA süsteem jagab salvestatud info kategooriatesse ning arvutab täiskasvanute sõnade arvu (ema, isa, hooldaja jne), lapse sõnade arvu ja lapse voorude arvu ning võimaldab infot näha graafikutena. Kogu päevane heli edastatakse lisaks protsentidena - elektroonilise heli, vaikuse osakaal, sisuline kõne jne. Vooruvahetuseks registreeritakse, kui täiskasvanu ütleb midagi ja laps reageerib või vastupidi [www.lena.org].

Lisaks saab uuritavaid katseisikuid võrrelda standardiseeritud skooridega, mis on ingliskeelse valimi põhjal välja töötatud 2–48 kuu vanustele lastele (Gilkerson & Richards, 2008).

***ECDI-II*** – (*MacArthur–Bates Communicative Development Inventory: Words and Sentences eestikeelne adaptatsioon*)

ECDI-II mõõdab 16–30 kuu vanustel lastel aktiivset sõnavara, grammatikareeglite valdamist ning seda, kas laps oskab rääkida olevikust, minevikust ja asjadest, mida kohal ei ole. Test ECDI-II on adapteeritud eesti keelde ja sisaldab 631 sõna, mis on jaotatud 21 kategooriasse (Schults, Tulviste, Haan, 2013; Tulviste, 2007; Urm & Tulviste, 2016), näiteks: loomad, riideesemed, häämitsused, ajamäärused, asesõnad, hulgasõnad jm. Test täidetakse lapsevanema poolt oma lapse kõnearengu kohta.

***CLAN*** (*Computerized Language Analysis*) on arvutipõhine keeleanalüüsi süsteem. Käesolevas töös transkribeeriti LENA seadmega salvestatud audiofailide väljavalitud lõigud kasutades CHILDES süsteemi (MacWhinney, 2019). Transkribeerimiseks kasutati CHAT (Codes for the Human Analyses of Transcripts) formaati ning transkriptsioone analüüsiti CLAN programmis.

Vaadeldi lapse kõne hulka (öeldud sõnade arv), vanema kõne hulka ning lapse voorude hulka. CLAN MLT programm võimaldab kokku lugeda sõnu ja vooruvahetusi ning arvutab lausete arvu voores, sõnade arvu lauses ja sõnade arvu voores kohta.

Voor on defineeritud CLAN-i instruksiooni järgi, kui ühe kõneleja poolt räägitavate lausete jada. Kui sama kõneleja järjest räägib, on iga lausung osa käesolevast voores ning need arvutused esitatakse iga kõneleja kohta eraldi (MacWhinney, 2019).

## Tulemused

Andmeanalüüsi meetoditest kasutati erinevaid kirjeldava statistika meetodeid, korrelatsioonianalüüsi (täpsemalt Spearmani mitteparameetriline korrelatsioonianalüüs), mitteparameetriline Mann–Whitney U–test ning Kruskal–Wallise test.

Andmeanalüüs teostati paketiga SPSS (23. versioon).

### I osa

Käesoleva töö esimeses osas uuriti, kui täpselt LENA kõneanalüüsi süsteem tuvastab laste vokalisatsioonide, vanemate sõnade ja voorude hulga eesti keeles eesmärgiga välja selgitada, kas LENA seade loeb sõnu ja voore sama usaldusväärselt nagu inimene. Kasutati korrelatsioonianalüüsi, valimi väiksuse tõttu (mõõdetud tunnused ei allunud normaaljaotusele) valiti selleks Spearmani korrelatsioonikoefitsient.

**Uurijatevahelise reliaabluse** väljaselgitamiseks transkribeeriti teise eesti keelt emakeelena kõneleva isiku poolt 10% transkribeeritud lõikude kogupikkusest. Transkribeerijate erinevuse hindamiseks (vt Lisa 1) leiti erinevus protsentides ning suhtarv (teise transkribeerija (TR2) tulemuste suhe esimese transkribeerija (TR1) tulemustesse).

Kahe transkribeerija erinevus täiskasvanute sõnade puhul suhtarvuna tuli 1,2 (erinevus +14,6%), laste vokalisatsioonide puhul oli suhtarv 0,9 (erinevus -16,3%) ja laste voorude puhul 0,8 (erinevus -28,2%). Sõnade, vokalisatsioonide ja voorude arvestuses kokku tuli erinevus kahe transkribeerija vahel suhtarvuna 1,02 (erinevus +1,7%). Seega olid kahe transkribeerija vokalisatsioonide, sõnade ja voorude loendites arvestatavad erinevused, ent kumbki transkribeerija ei lugenud sõnu kolmes kategoorias läbivalt rohkem või vähem kui teine. Transkribeerija TR2 luges vähem täiskasvanute sõnu ja voore ning transkribeerija TR1 luges vähem laste vokalisatsioone.

### LENA kõneanalüüsi süsteemi usaldusväärsus

LENA reliaabluse väljaselgitamiseks analüüsiti 10 minutilisi lõike iga uuritava kohta, kokku 220 minutit (LENA vs inimene). Kõik analüüsid on samuti tehtud 22 uuritava kohta.

10-minutilises lõigus transkribeerija loendatud laste vokalisatsioonide ja LENA poolt loendatud laste vokalisatsioonide vahel ilmnes tugev seos ( $r=.72$ ,  $p<.01$ ) (Tabel 2). Transkribeerija ja LENA loendatud täiskasvanu sõnade puhul ilmnes samuti statistiliselt oluline positiivne seos ( $r=.75$ ,  $p<.01$ ). Statistiliselt oluline mõõdukas seos avaldus transkribeerija ja LENA loendatud lapse voorude vahel ( $r=.61$ ,  $p<.01$ ).

Korrelatsioonianalüüs kinnitas, et LENA kõneanalüüsi süsteem loeb kõigis kolmes segmendis (laste vokalisatsioonid, laste voorud, täiskasvanute sõnad) sarnaselt transkribeerijale ning annab sarnaseid tulemusi.

**Tabel 2.** 10-minutiliste lõikude LENA loenduste ja inimtranskribeerija loenduste seosed kogu valimil

	1	2	3	4	5	6
1.LENA laste vokal-d	-					
2.LENA täiskasv sõnad	-.32	-				
3.LENA laste voorud	-.10	<b>.84**</b>	-			
4.Transkrib* laste vokal	<b>.72**</b>	-.17	.09	-		
5.Transkrib* laste voorud	-.24	<b>.61**</b>	<b>.61**</b>	.09	-	
6.Transkrib* täiskasv sõnad	-.33	<b>.75**</b>	<b>.61**</b>	-.08	<b>.65**</b>	-

\*Transkrib – transkribeerija poolt loendatud sõnade arv

Statistiliselt olulised seosed \*\* $p < .01$

### Kahe meetodi võrdlus

Lisaks viidi läbi kahe meetodi (inimene vs LENA) võrdlev analüüs. Selleks loodi uued tunnused, mis väljendavad kahe meetodi rakendamisel ilmnunud erinevusi esiteks absoluutarvudena ning teiseks osakaaluna protsentides iga indiviidi puhul. Kuna eelnevate uuringute kohaselt kaldub LENA lugema vähem sõnu kui inimene, lahutati transkribeeritud ja CLANi poolt kokku loendatud sõnade arvust LENA sõnade arv (vt Tabel 3). Erinevused arvutati nii laste vokalisatsioonide, täiskasvanute sõnade kui ka laste vooruvahetuste puhul.

Laste vokalisatsioonide puhul tuli 10-minutilise lõigu keskmiseks erinevuseks 172 sõna (minimaalne erinevus 7 ja maksimaalne 321 sõna; (SD=92). LENA luges inimese transkriptsioonidega võrreldes laste vokalisatsioonide puhul oodatult vähem sõnu: kahe erineva mõõtmise tulemuse keskmine protsentuaalne erinevus oli 50% (SD=18). Võrreldes kahe teise segmendiga on erinevus just siin kõige suurem.

Täiskasvanute sõnade puhul tuli võrreldud LENA ja transkriptsioonide 10-minutilise lõigu keskmiseks erinevuseks 42 sõna (minimaalne erinevus -167, maksimaalne 396; SD=112). LENA loendatud sõnade hulk oli suurem kui transkribeerijal, protsentides oli vastav erinevus 6% (SD=73).

Laste vooruvahetuste puhul kahe meetodi võrdluses tuli keskmine erinevus 19 vooru (minimaalselt 11, maksimaalselt 52; SD=13). LENA luges siinkohal samuti nagu laste sõnade korral, vähem vooruvahetusi kui inimene, protsentides oli erinevus 48% (SD=31).

**Tabel 3.** Kahe meetodi 10-minutiliste lõikude erinevused absoluutarvudes ja protsentides (N=22)

	Min	Max	Keskmine	SD	Min %	Max%	Keskmine%	SD%
Laste sõnad	7	321	172	92	7	67	50	18
*TK sõnad	-167	396	42	112	-227	75	-6	73
Voorud	-11	52	19	13	-44	96	48	31

\*TK – täiskasvanute sõnad

## II osa

### Laste keelekeskkonna analüüs

Esmalt uuriti kirjeldava statistikaga kõne hulka, mida lapsed ühel salvestuspäeval kuulsid ja ise produtseerisid (kõik andmed olid kogutud 22 uuringus osaleja kohta). Töö järgnevates analüüsides kasutati LENA kõneanalüüsi süsteemi loendustulemusi laste vokalisatsioonide, vanemate sõnade ja laste voorude osas kogu salvestuspäeva kohta.

Vanemate sõnade hulk ühe salvestuspäeva jooksul oli väga erinev. Kõige rohkem kuulis laps ühe päeva jooksul 32 381 sõna (vt Tabel 4), kõige vähem 9 788 sõna (M=19 591; SD=5 352).

Laste voorude hulk ühe salvestuspäeva jooksul oli LENA andmeil maksimaalselt 1 526 lapse vooru ja minimaalselt 429 vooru päevas (M=917; SD=283).

Laste vokalisatsioone uuritud päeval oli maksimaalselt 5 769 ja minimaalselt 2 123 vokalisatsiooni päevas (M=3 724; SD=1 078).

**Tabel 4.** Laste vokalisatsioonide, täiskasvanute sõnade, laste voorude hulk salvestuspäeva jooksul (N=22)

	Miimum	Maksimum	Keskmine	SD
Laste vokal-d	2123	5769	3724	1078
TK sõnad*	9788	32381	19592	5352
Voorud	429	1526	917	283

\*TK – täiskasvanute sõnad

### Uuringugrupi ECDI–II testi tulemuste võrdlus üldpopulatsiooniga

Laste ECDI–II testi tulemusi võrreldi 100 3-aastase lapse testitulemustega (vt Tabel 5). Uuringugruppi kuuluvatest lastest vastas poolte laste tulemus 50 protsentiilile. Kahe lapse



tulemused jäid 10 protsentiili piirile ja 3 lapse tulemus 90 protsentiili piirile. Tulemused vastasid üldisele keskmisele ehk ei olnud üliheade tulemustega või väga madalate tulemustega lapsi.

Kuigi käesolev valim oli väga tagasihoidliku suurusega (N=22), võib väita, et ECDI–II skoori alusel vastab uuritav valim üldjoontes üldpopulatsioonile.

**Tabel 5.** Uuringugrupi võrdlus 100 3-aastase lapse ECDI–II (T. Tulviste publitseerimata andmed) testi tulemustega

Protsentiilid (N=100)	10%	25%	50%	75%	90%
Keskmine sõnade arv (N=100)	169	384,5	443	569,5	605
<b>Käesolev uuring (N=22)</b>	<b>2</b>	<b>2</b>	<b>11</b>	<b>4</b>	<b>3</b>

#### Poiste ja tüdrukute tulemuste erinevused (LENA loendused ja ECDI–II tulemused)

Tüdrukute kõne areng kaldub olema varasemate uuringute andmetel poiste omast kiirem (Yu jt, 2014; Eriksson jt, 2012), seetõttu sooviti välja selgitada, kas erinevused esinevad ka uuritaval valimil sugude lõikes. Selleks vaadeldi laste poolt produtseeritud vokalisatsioonide, voorude hulka ning kõne hulka, mida nad ümbritsetavatelt täiskasvanutelt kuulevad ja nende ECDI–II testi tulemusi.

Poiste ja tüdrukute ECDI–II testitulemuste keskmiste võrdlustes oli poiste keskmine punktisumma 506 (SD=94) ja tüdrukutel 521 (SD=69) (vt Tabel 6). Tüdrukutel olid testitulemused keskmiselt mõneti kõrgemad.

Salvestuspäeval rääkisid vanemad tüdrukutega rohkem kui poistega. Vanemad rääkisid poistega salvestuspäeva jooksul LENA loenduste järgi keskmiselt 18 692 (SD=7 551) sõna ja tüdrukutega 20 214 (SD=3 330) sõna.

Vooruvahetusi oli LENA loenduste järgi poistel keskmiselt 871 (SD=305) ja tüdrukutel keskmiselt 949 (SD=275).

Laste vokalisatsioonide järgi oli jaotus soo lõikes LENA järgi keskmiselt 3 505 (SD=954) sõna poistel ja 3 877 (SD=1 168) sõna tüdrukutel.

Soolise erinevuse statistilise olulisuse hindamiseks on tavapärase t–testi asemel kasutatud mitteparameetrilist Mann–Whitney U–testi. Põhjuseks on valimi väga väike maht ning andmete mittevastavus normaaljaotusele.

Mitte ühegi eelpool mainitud tunnuse puhul ei osutunud erinevus statistiliselt oluliseks (vanemate sõnade ja puhul  $p=.16$ ; laste voorude puhul oli  $p=.36$ ; laste vokalisatsioonide puhul  $p=.51$  ning ECDI–II tulemuse puhul  $p=.70$ ).

**Tabel 6.** LENA loenduste ja ECDI–II testi tulemuste soolised erinevused (N=22)

	Sugu	N	Keskmine	SD	Mediaan
ECDI–II	Poiss	9	506	94	493
	Tüdruk	13	521	69	524
TK* sõnad	Poiss	9	18 692	7 551	16 725
	Tüdruk	13	20 214	3 330	20 356
Voorud	Poiss	9	871	305	848
	Tüdruk	13	949	275	965
Laste vokal-d	Poiss	9	3 505	954	3 280
	Tüdruk	13	3 877	1 168	3 696

\*TK – täiskasvanute sõnad

### **ECDI-II testi tulemuste ja LENA loenduste seosed üldvalimil ning sugude lõikes**

Järgnevate analüüsidega otsiti vastuseid teisele uurimisküsimusele kas ECDI–II testi tulemused ja LENA loendatav laste vokalisatsioonide, vooruvahetuste ning täiskasvanute sõnade arv on koondvalimil ja eraldi poiste ning tüdrukute alavalimil omavahel positiivselt seostud. Kasutati Spearmani mitteparameetrilist korrelatsioonianalüüsi, kuna uuritavate tunnuste jaotused ei allu normaaljaotusele ning valimi maht on tagasihoidlik.

Korrelatsioonikoeffitsiendi väärtused laste vokalisatsioonide, täiskasvanute sõnade ja laste voorude arvu ning ECDI–II testitulemuste vahel olid läbivalt positiivsed, kuid mitteolulised (vt Tabel 7).

**Tabel 7.** ECDI–II testi tulemuste ja LENA poolt loendatud sõnade ja voorude hulga seosed üldvalimil ning poiste ja tüdrukute alavalimil

	Üldvalim ECDI-II	Poisid ECDI-II	Tüdrukud ECDI-II
Täiskasvanu sõnad	.30	<b>.75*</b>	.02
Laste voorud	.14	.30	.03
Laste vokalisatsioonid	.06	-.17	.22

Statistiliselt oluline seos \* $p < .05$

LENA poolt loetud lastevanemate sõnade, laste voorude ning laste vokalisatsioonide arv ei ole ECDI–II testiga üldvalimil statistiliselt olulises seoses.

Lisaks viidi läbi korrelatsioonianalüüs ECDI–II testitulemuste ja LENA loendatud sõnade, voorude seoste uurimiseks soo lõikes eraldi poiste ja tüdrukute alavalimil.

Poiste alavalimi puhul torkab silma tugev positiivne seos ECDI–II testi tulemuste ja täiskasvanute sõnade arvu vahel ( $r = .75$ ,  $p = .02$ ). Ehk siis poiste puhul seostub suurem täiskasvanute loendatud sõnade arv kõrgema ECDI–II skooriga.

Tüdrukute alavalimi puhul ei ilmnenud ühtegi olulist korrelatsiooniseost.

### **Vanemate haridustaseme mõju laste ECDI–II testi tulemustele**

Mitmed varasemad uuringud on näidanud positiivseid seoseid vanemate hariduse ja lapse sõnavara suuruse vahel – kõrgema haridustasemega vanemate laste sõnavara on suurem. Selle uurimisküsimusele vastuse saamiseks jaotati uuritav valim kolme gruppi vanemate haridustaseme järgi: esiteks lapsed, kelle mõlemal vanemal oli kõrgharidus; teiseks lapsed, kelle üks vanematest omas kõrgharidust ning kolmandaks lapsed, kelle vanematel oli kesk- ja/või kutse- ja/või põhiharidus. Analüüsi kaasati 21. lapse andmed, kuna ühe uuritava lapse vanema hariduse kohta andmed puudusid.

**Tabel 8.** Laste ECDI–II testi punktisummade keskmised vanemate haridustasemete järgi

	N	Keskmine	SD
Mõlemad vanemal kõrgem haridus	8	516	86,6
Vähemalt ühel vanemal kõrgem haridus	10	501	71,5
Muu haridustase	3	541	107,9
Kokku	21	512	79,2

Nagu ülalolevast tabelist on näha (Tabel 8), siis kõrgharidusega vanemate lastel ei ole keskmine ECDI–II skoor kõrgem muu haridustasemega vanemate laste grupist. Kuna valim on väike, siis ei kinnitanud Kruskal–Wallise test gruppide erinevust statistilisel olulisel tasemel ( $p = .67$ ).

Lisaks võrreldi kõrgharidusega emade laste keskmist skoori ülejäänud laste keskmise skooriga (Tabel 9) ning tulemuseks oli väga tagasihoidlik erinevus: kõrgema haridustasemega emade laste punktisumma oli 515 ning muu haridustasemega emade laste puhul 513 punkti. Ootuspäraselt Mann–Whitney U–test statistilist erinevuste ei kinnitanud ( $p>.05$ ).

**Tabel 9.** Laste ECDI–II testitulemuste keskmised ema haridustaseme järgi

	N	Keskmine	SD
Kõrgem haridus	16	515	77,1
Muu haridus	6	513	88,3
Kokku	22	514	78,1

## Arutelu

LENA tarkvara loodi ingliskeelse keskkonna keeleuuringute tarbeks, kuid selle usaldusväärsus on varasemates uuringutes kinnitust leidnud ka mitmetes mitte-inglisekeelsetes keskkondades. Eesti keeleruumis ei ole LENA kõneanalüüsi süsteemi laste keelekeskkonna uurimiseks varem kasutatud. Uuring toimus suurema longituuduuringu „Varase kasvukeskkonna toetav roll lapse kõneoskuste arengus: võimalus kahandada lõhet hilisemates akadeemilistes saavutustes“ (PUT 1359, vastutav täitja Tiia Tulviste) raames.

Käesoleva töö esimese osa eesmärk oli välja selgitada, kas LENA kõneanalüüsi süsteemi kasutamine eesti keelekeskkonnas annab usaldusväärseid tulemusi ning kui suur on LENA poolt loendatud sõnade ja voorude erinevus laste ning täiskasvanute puhul võrreldes inimese tehtud transkribeeringutega.

LENA kõneanalüüsi süsteemi usaldusväärsuse väljaselgitamiseks ja esimesele uurimusküsimusele vastuse leidmiseks viidi käesoleva uurimuse raames läbi korrelatsioonianalüüs transkribeerija ja LENA loenduste vahel kolmes osas: laste vokalisatsioonid, laste voorud, täiskasvanute sõnad. LENA poolt loendatud laste vokalisatsioonide ja transkribeerija poolt kirja pandud sõnade võrdluses ilmnis tugev seos, mis oli sarnane varasemate hollandlaste (Busch jt, 2017) ja prantslaste (Canault jt, 2015) uuringute tulemustega. Statistiliselt olulised olid käesolevas uuringus ka seosed transkribeerija ja LENA loendatud täiskasvanu sõnade vahel, mille puhul saadi tugev seos, mõõdukas seos saadi laste voorude puhul. Võrreldes uuringutega rootsi ning prantsuse keeltes saadi käesolevas töös täiskasvanute sõnade puhul kahe loendusmeetodi tulemustes madalamaid seoseid (Schwarz jt, 2017; Canault jt, 2015). Voorude puhul saadi Hollandi uuringus inimese ja LENA loendustulemuste võrdlemisel madalam seos (Busch jt, 2017), Hiina ja Vietnami uuringus tugevam seos võrreldes käesoleva töö tulemusega (Gilkerson jt, 2015; Ganek & Eriks-Brophy, 2018).

Käesoleva töö korrelatsioonianalüüs kinnitas, et LENA kõneanalüüsi süsteem loeb eesti keeles kõigis kolmes segmendis (laste vokalisatsioonid, laste voorud, täiskasvanute sõnad) sarnaselt transkribeerijale, ent vaja oleks täiendavaid täpsustavaid uuringuid kõneanalüsaatori kliiniliseks kasutamiseks eesti keelekeskkonnas.

LENA kõneanalüüsi süsteemi ja inimese transkribeeringute erinevuse väljaselgitamiseks võrreldi neid kahte meetodit. Tulemusena selgus, et erinevus oli kõige suurem laste sõnade arvu lugemisel. LENA süsteem luges keskmiselt 50% vähem laste sõnu kui inimene. LENA süsteem

luges näiteks ka Hollandi ja Prantsuse uuringutes vähem sõnu kui transkribeerija (Busch jt, 2017; Canault jt, 2015). Täiskasvanute sõnade puhul tuli käesolevas töös võrreldud LENA ja transkriptsioonide 10-minutilise lõigu keskmiseks erinevuseks 6%, kusjuures LENA loendatud sõnade hulk oli suurem kui transkribeerijal. Sarnase tulemuse said ka Gilkerson jt (2015), kelle uuringus luges LENA samuti rohkem täiskasvanu sõnu kui transkribeerija. Laste vooruvahetuste puhul luges LENA 48% voorudest vähem vooruvahetusi kui inimene. Sarnane erinevus on varem avaldunud ka näiteks Busch'i ja tema kaastöötajate (2017) töös. Erinevused kahe meetodiga saadud laste kõnet puudutavate loendustulemuste vahel olid käesolevas töös väga suured. Kahe meetodi erinevuse väljaselgitamine eesti keeles vajaks täiendavaid uuringuid suuremal valimil täpsemate transkribeerimiskriteeriumitega. Varasemate uuringute puhul on samuti välja toodud kahe meetodi loendustulemuste suuri erinevusi ning püütud neid erineval viisil põhjendada. Ühe põhjusena on Elo (2016) viidanud keeleliste eripäradele. Soome keeles on palju liitsõnu, mis võib põhjustada sõnade valesti loendamist LENA poolt. See põhjendus võib kõne alla tulla ka eesti keeles, ehkki näiteks täiskasvanute sõnade puhul see erinevus käesolevas uuringus ei ilmnenu – LENA kõneanalüüsi süsteem luges täiskasvanute sõnu rohkem kui inimene. Lisaks on uurijad loenduste erinevuste põhjendustena välja toonud uuringute erineva metoodika (Busch jt, 2017). Laste vokalisatsioonide lugemisel tekkivate erinevuste põhjendusena on konkreetselt välja pakutud õues salvestamisega kaasnevat müra ja riidetuse sahinat (Canault jt, 2015). Käesoleva töö autor täheldas raskusi kahe sarnases vanuses lapse jutu transkribeerimisel mängusituatsioonis. Laste kõne eristamine on keeruline, kui lapsed on mängu ajal üksteisele väga lähedal.

Töö esimese osas selgitati lisaks välja kahe transkribeerija erinevus. Täiskasvanute sõnade puhul tuli erinevus suhtarvuna 15%, lapse vokalisatsioonide puhul 14% ja laste voorude puhul 22%. Sõnade ja voorude arvestuses kokku tuli erinevus 2%. Seega, ehkki erinevused olid arvestatavad, ei olnud transkribeerijate tulemustes püsivat kallutatust: üks transkribeerija ei lugenud sõnu pidevalt rohkem või vähem kui teine. Mõningates varasemates töödes on saadud transkribeerijate vahel väga kõrget kokkulangevust ( $r=.99$ ,  $p<.001$ ; Canault jt, 2015 või  $r=.95$ ,  $p<.01$ ; Schwarz jt, 2017). Ent nendes töödes on uurijate poolt kirjeldatud transkribeerijate eelnevat väljaõpet transkribeeringute ühtlustamiseks, mida meie uuringu puhul ei olnud. Võrreldava materjali suurem hulk annaks samuti täpsema tulemuse.

Töö teises osas oli eesmärgiks uurida ECDI-II testi tulemuste ja LENA loenduste seoseid uuringugrupi üldvalimil ning sugude lõikes vastavalt püstitatud uurimisküsimustele. Lisaks uuriti seda, kui palju kõnet kuulsid uuritavad lapsed täiskasvanutelt ühe salvestuspäeva jooksul

ja kui palju lapsed ise rääkisid. Viimase küsimusena oli uurimise all, kuidas on seotud vanemate haridustase ja laste ECDI–II testi tulemused. Uuritavate laste ECDI–II testi tulemusi võrreldi 100 3–aastase lapse sama testi tulemustega ning nende testide tulemused vastasid üldisele keskmisele ehk ei olnud üliheade tulemustega või väga madalate tulemustega lapsi.

Analüüside tulemusel ilmnedid nõrgad positiivsed korrelatsioonid laste vokalisatsioonide arvu, täiskasvanute sõnade ja laste voorude ning ECDI–II testi tulemuste vahel. Need seosed osutusid üldvalimil mitteolulisteks. Varasemalt on Elo (2016) leidnud positiivse seose 24 kuu vanuste laste MC–CDI testi ja täiskasvanute sõnade skoori vahel, Warreni tööühm (2009) aga täiskasvanute sõnade, lapse voorude ja CDI vahel.

Käesoleva töö puhul võisid mainitud tulemused olla põhjustatud asjaolust, et üks salvestuspäev lapse peres on ebapiisav. Näiteks võis üks vanem viibida salvestuspäeval kodust eemal ja laps ei kuulnud nii palju kõnet kui tavaliselt, ning lapsevanema sõnade hulga mõju lapse kõne hulgale ei avaldunud tulemustes, mõjutades ka lapse sõnade ja voorude hulka. Osades varasemates töödes on andmeid kogutud mitmepäevaste salvestustena (Canault jt, 2015; Weisleder & Fernald, 2013) ning see võib anda täpsema ettekujutuse vanema ja lapse suhtlemise mõjust lapse sõnavara arengule.

Lisaks viidi käesolevas töös läbi korrelatsioonianalüüs ECDI–II testitulemuste ja LENA loendatud sõnade ja voorude seoste uurimiseks poiste ja tüdrukute vahel. Selle analüüsi oluliseks tulemuseks oli tugev positiivne seos täiskasvanute sõnade ja poiste ECDI–II testi tulemuste vahel ( $r=.75$ ,  $p=.02$ ). Poiste tulemus räägib selle kasuks, et kui täiskasvanud lastega rohkem kõnelevad, on laste kõne areng kiirem, toetades kõne arengu sotsiaal-interaktsioonilist teooriat (Hart & Risley, 2003; Hoff, 2006). See on käesoleva töö puhul tähelepanuväärne ja muude tulemuste seast esile kerkiv leid, mis näib kinnitavat oletust keskkonna mõju kohta laste keelelisele arengule. Varasemates uuringutes on täheldatud, et poiste kõne areng kaldub olema tüdrukute kõne arengust aeglasem, poisid räägivad vähem ja nendega räägitakse vähem (Yu jt, 2014; Eriksson jt, 2012), seega võib pakkuda välja oletuse, et kui poistega rohkem rääkida võib nende kõne areng olla kiirem. Leitud tulemus vajaks kahtlemata edaspidi uurimist suuremal valimil usaldusväärsema tulemuse saamiseks. Tüdrukute valimil käesolevas töös olulisi seoseid ei ilmnunud.

Töö teises osas uuriti lisaks kirjeldava statistikaga kõne hulka, mida lapsed ühel salvestuspäeval kodus kuulsid ja ise produtseerisid. Valimisse kuulunud laste ühe päeva jooksul kuuldud sõnade arv oli suure erinevusega – 22 593 sõna. Suur erinevus kuuldud sõnade hulgas vastab varasematele uuringutele. Näiteks leidsid Weisleder ja Fernald (2013) 19–24 kuu vanuseid

hispaaniakeelseid lapsi uurides, et 10 tunni jooksul jäi laste kuulnud kõne hulk vahemikku 2 000 kuni 29 000 sõna, seega vahe oli 27 000 sõna. Laste vokalisatsioonide puhul oli vahe käesolevas uuringus 3 646 sõna ning laste voorude hulga vahe ühe salvestuspäeva jooksul oli LENA andmeil 1 097 vooru.

Täiendavalt uuriti soolisi erinevusi laste kõne arengus laste sõnade arvu, vanemate sõnade arvu, laste vokalisatsioonide ja ECDI–II testitulemuste osas. Kõigis neljas võrreldavas kategoorias olid tüdrukute tulemused paremad kui poistel. Need soolised erinevused (LENA loendatud sõnade arvus, voorudes ja sõnavara suurused) ei olnud aga statistiliselt olulised ilmselt valimi väiksuse tõttu.

Võrreldes erinevaid vanemate hariduse tasemeid ja laste ECDI–II testi tulemusi ootasime arvestatavaid erinevusi laste sõnavara suuruses, nii nagu mitmetes varasemates töödes on leitud (Romeo jt, 2018; Schulsi jt, 2012; Hoff-Ginsberg, 1998). Statistilisi erinevusi erineva haridustasemega emade laste sõnavaras antud töös ei ilmnenud. Lastel, kelle emal oli kõrgharidus oli vaid pisut suurem ECDI–II testiga mõõdetud sõnavara kui põhi-, kutse- ja/või keskharidusega vanemate lastel.

Uuringu kitsaskohana võib välja tuua valimi väiksuse, kuna andmete kogumine oli piiratud ühe LENA salvestusseadmega. Käesolevas töös püstitatud uurimisküsimused vajaksid uurimist suuremal valimil. Edasiste uuringute puhul võib kaaluda salvestamisi mitme päeva jooksul, mis annaksid üldistavamalt infot lapse keelekeskkonna kohta ning võimaldaks usaldusväärsemaid tulemusi. Tulemuste usaldusväärsuse huvides oleks edaspidi oluline tähelepanu pöörata ka transkribeerijate väljaõppele salvestiste ümberkirjutuste ühtlustamiseks. Võib kaaluda ka uurijate tööd kiirendavate programmide kasutamist (näiteks Praat vms), et kulutada vähem inimressurssi ja aega.

Lõppvalim jäi maksimaalsest võimalikust (25 last) pisut väiksemaks, kuna sügis-talvisel perioodil haigestusid lapsed üsna tihti viirushaigustesse, mistõttu salvestamised lükkusid edasi ja pikenes ka andmete kogumise periood. Valimi suurendamisel said takistuseks ka suviselt soojad ilmad, mil lapsed viibisid rohkem õues ning veekogude ääres ja salvestusseadme kandmine oli kas vastavalt tülikas või isegi ebasoovitav.

#### Kokkuvõte

Käesolev uuring oli esmane katsetus LENA kõneanalüüsi süsteemi kasutamiseks eesti keeleruumis. Uuringu tulemused annavad aluse täpsemateks ja mahukamateks uuringuteks. Uuringu peamise tulemusena selgus, et LENA kõneanalüüsi süsteemi saab kasutada eesti laste



keelekeskkonna uurimisel. Kuna erinevused LENA loenduste ja inimese transkriptsioonide vahel olid suured, peaks edaspidistes uuringutes kaaluma kõneanalüüsi süsteemi rakendusvaldkondi (kuidas ja kus süsteemi kasutada) ning täpsustavalt uurima süsteemi usaldusväärsust. Seoste uurimisel ECDI–II testitulemuste ja LENA loenduste vahel ilmnas poiste valimil täiskasvanute sõnade ja testitulemuste puhul tugev positiivne seos, mis näitab, et just poiste ekspressiivse sõnavara arengu seisukohalt on oluline poistega rohkem suhelda. Üldvalimil ning tütarlaste alavalimil seoseid ECDI–II testi tulemuste ja LENA loenduste vahel ei õnnestunud kinnitada. Samuti selgus uurimusest, et vanemate kõrgem haridustase ei seostunud uuritud valimi puhul laste ECDI–II testi tulemustega. Lastel, kelle emal oli kõrgharidus ilmnas veidi suurem ECDI–II testi keskmine punktisumma, ent tulemus ei osutunud statistiliselt oluliseks.

LENA kõneanalüüsi süsteemi kasutuselevõtt hõlbustaks meie keeleteadlaste ja psühholoogide tööd kõne uuringute läbiviimisel, lihtsustaks klinitistide tööd laste arenguprobleemide väljaselgitamisel ning oleks abiks sekkumiste pakkumisel lastele ja nende vanematele.

### **Tänu sõnad**

Täna südamest kõiki toremaid perekondi ja armsaid lapsi, kes sellesse uuringusse olid kaasatud, samuti juhendajaid, sõpru ning oma perekonda toetuse ja abi eest töö valmimisel.

### Kirjanduse loetelu

- Berglund, E., Eriksson, M. & Westerlund, M. (2005). Communicative skills in relation to gender, birth order, childcare and socioeconomic status in 18-month-old children. *Scandinavian Journal of Psychology*, 46, 485–491
- Busch, T., Sangen, A., Vanpoucke, F., van Wieringen, A. (2017). Correlation and agreement between Language ENvironment Analysis (LENA™) and manual transcription for Dutch natural language recordings. *Behavior Research Methods*, 50(5), 1921-1932
- Campbell, T. F., Dollaghan, C. A., Rockette, H. E., Paradise, J. L., Feldman, H. M., Shriberg, L. D., Sabo, D. L., Kurs-Lasky, M. (2003). Risk factors for speech delay of unknown origin in 3-year-old children. *Child Development*, 74(2), 346-357
- Canault, M., Le Normand, M. T., Foudil, S., Loundon, N., & Thai-Van, H. (2015). Reliability of the Language ENvironment Analysis system (LENA™) in European French. *Behavior Research Method*, 48(3), 1109–1124
- Choudhury, N., Benasich, A. A. (2003). A family aggregation study: The influence of family history and other risk factors on language development. *Journal of Speech Language and Hearing Research*, 46(2), 261-272
- Elo, H. (2016). Acquiring language as a twin. Twin children's early health, social environment and emerging language skills. Academic Dissertation. *University of Tampere*. Retrieved from <https://tampub.uta.fi/bitstream/handle/10024/100187/978-952-03-0296-2.pdf?sequence=1&isAllowed=y>
- Eriksson, M., Marschik, P. B., Tulviste, T., Almgren, M., Perez Pereira, M., Wehberg, S., Gallego, C. (2012). Differences between girls and boys in emerging language skills: Evidence from 10 language communities. *British Journal Of Developmental Psychology*, 30, 326-343
- Ganek, H. & EriksBrophy, A. (2016). The Language ENvironment Analysis (LENA) System: A literature review. *Department of Speech-Language Pathology University of Toronto*. Retrieved from <https://www.ep.liu.se/ecp/130/004/ecp16130004.pdf>

- Ganek, H. V. & Eriks-Brophy, A. (2018). A concise protocol for the validation of Language ENvironment Analysis (LENA) conversational turn counts in Vietnamese. *Communication Disorders Quarterly*, 39(2), 371-380
- Gilkerson, J., Zhang, Y., Xu, D., Richards, J. A., Xu, X., Jiang, F., Topping, K. (2015). Evaluating LENA System performance for Chinese: A pilot study in Shanghai. *Journal of Speech, Language, and Hearing Research*, 58(2), 445-452
- Gilkerson, J. & Richards, J. (2008). The LENA natural language study (Technical Report LTR02-2). Boulder, CO: LENA Foundation. Retrieved from [https://www.lena.org/wp-content/uploads/2016/07/LTR-02-2\\_Natural\\_Language\\_Study.pdf](https://www.lena.org/wp-content/uploads/2016/07/LTR-02-2_Natural_Language_Study.pdf)
- Golinkoff, R., M., Hoff, E., Rowe, M. L., Tamis-LeMonda, C. S., Hirsh-Pasek, K. (2018). Language matters: Denying the existence of the 30-million-word gap has serious consequences. *Child Development*, 00(0), 1–8
- Hart, B., Risley, T. R. (2003). The early catastrophe. The 30 million word gap. *American Educator*, 27(1), 4-9
- Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, 74(5), 1368-1378
- Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, 26, 55-88
- Hoff-Ginsberg, E. (1998). The relation of birth order and socioeconomic status to children's language experience and language development. *Applied Psycholinguistics*, 19(4), 603-629
- MacWhinney, B. (2019). Tools for analyzing talk part 2: The CLAN Program. Brian Carnegie Mellon University. Retrieved from <https://talkbank.org/manuals/CLAN.pdf>
- Nicholas, J.,G & Ann E. Geers, A., E. (2006). Effects of early auditory experience on the spoken language of deaf children at 3 years of age. *Ear Hear*, 27(3), 286–298
- Paavola, L., Kunnari, S., Moilanen, I. (2005). Maternal responsiveness and infant intentional communication: Implications for the early communicative and linguistic development. *Child: Care, Health & Development*, 31, 727 – 735

- Pae, S., Yoon, H., Seol, A., Gilkerson, J., Richards, J., Ma, L., & Topping, K. (2016). Effects of feedback on parent-child language with infants and toddlers in Korea. *First Language*, 36(6), 549-569
- Romeo, R. R., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Rowe, M. L., Gabrieli, J. D. (2018). Beyond the 30-million-word gap: Children's conversational exposure is associated with language-related brain function. *Psychological Science*, 29(5), 700-710
- Schults, A., Tulviste, T., Haan, E. (2013). Early vocabulary in full term and preterm Estonian children. *Early Human Development*, 89(9), 721 – 726
- Schults, A., Tulviste, T., & Konstabel, K. (2012). Early vocabulary and gestures in Estonian children. *Journal of Child Language*, 39, 664–686
- Schwarz, I.-C., Botros, N., Lord, A., Marcusson, A., Tideliu, H., & Marklund, E. (2017). The LENA system applied to Swedish: Reliability of the adult word count estimate. *The International Speech Communication Association (ISCA)*. Retrieved from
- Sperry, D. E., Sperry, L. L., Miller, P. J. (2018). Language does matter: But there is more to language than vocabulary and direct speech. *Child Development*, 00(0), 1-5
- Stromswold, K. (2001). The heritability of language: A review and metaanalysis of twin, adopted, and linkage studies. *Language*, 77(4), 647-723
- Zimmerman, F. J., Gilkerson, J., Richards, J. A., Christakis, D. A., Xu, D., Gray, S., & Yapanel, U. (2009). Teaching by listening: The importance of adult-child conversations to language development. *Pediatrics*, 124(1), 342-349
- Tulviste, T. (2007). Variation in vocabulary development among Estonian children as a function of child's gender, birth order, child-care, and parental education. *Proceedings from the First European Network Meeting on the Communicative Development Inventories* (pp. 16-21). University of Gävle
- Urm, A., Tulviste, T. (2016). Sources of individual variation in Estonian toddlers' expressive vocabulary. *First Language*, 36(6), 580-600
- VanHulle, C. A., Goldsmith, H. H. & Lemery, K. S. (2004). Genetic, environmental and

gender effects on individual differences in toddler expressive language. *Journal of Speech, Language, and Hearing Research*, 47, 904-912

Warren, S. F., Gilkerson, J., Richards, J. A., Oller, K. D., Xu, D., Yapanel, U., & Gray, S.

(2010). What automated vocal analysis reveals about the vocal production and language learning environment of young children with autism. *Journal of Autism and Developmental Disorders*, 40(5), 555-569

Weisleder, A., & Fernald, A. (2013), Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, 24(11), 2143-2152

Westerlund, M. & Lagerberg, D. (2007). Expressive vocabulary in 18-month-old children in relation to demographic factors, mother and child characteristics, communication style and shared reading. *Child: Care, Health and Development*, 34(2), 257–266

Yu, V. Y., MacDonald, M. J., Oh, A., Hua, G. N., De Nil, L. F., Pang, E. W. (2014). Age-related sex differences in language lateralization: A magnetoencephalography study in children. *Developmental Psychology*, 50(9), 2276-2284

<https://www.lena.org/>

**Lisa 1.** Kahe transkribeerija (TR1 ja TR) võrdlus täiskasvanute sõnade, laste sõnade ja voorude loendamisel

	Täiskasvanu sõnad	Laste vokalisatsioonid	Laste voorud	Kokku
TR1	1045	566	110	1721
TR2	893	658	141	1692
Suhtarv TR1-TR2	1,2	0,9	0,8	1,02
Erinevuse %	14,6%	-16,3%	-28,2%	1,7%

## **Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks**

Mina, Jane Kuperjanov,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose “Lapse sõnavara ja tema keelekeskkond: LENA kõneanalüüsi süsteemi kasutamine eesti laste keelekeskkonna mõõtmisel”, mille juhendajad on Tiia Tulviste ja Ada Urm;

1.1. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

1.2. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile;

3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus 27.05.2019